Do as AI say: susceptibility in deployment of clinical decision-aids

Abstract

Artificial intelligence (AI) models for decision support have been developed for clinical settings such as radiology, but little work evaluates the potential impact of such systems. In this study, physicians received chest X-rays and diagnostic advice, some of which was inaccurate, and were asked to evaluate advice quality and make diagnoses. All advice was generated by human experts, but some was labeled as coming from an AI system. As a group, radiologists rated advice as lower quality when it appeared to come from an AI system; physicians with less task-expertise did not. Diagnostic accuracy was significantly worse when participants received inaccurate advice, regardless of the purported source. This work raises important considerations for how advice, AI and non-AI, should be deployed in clinical environments.